European Geosciences Union
General Assembly 2011

Vienna, Austria,
03-08 April 2011

Session: HS7.5/NP6.7
Hydroclimatic stochastics

Entropy maximization, *p*-moments and power-type distributions in nature

S.M. Papalexiou and D. Koutsoyiannis

Department of Water Resources and Environmental Engineering,
National Technical University of Athens, Greece (www.itia.ntua.gr)

Contact Info: smp@itia.ntua.gr

Available online: http://itia.ntua.gr/1127

1. Abstract

Choosing a proper probabilistic model for geophysical processes is not a trivial task. The common practice of choosing one of a few popular (among infinitely many) distributions is subjective and relies too much on empirical considerations e.g., the summary statistics of the data record. In contrast, the principle of maximum entropy offers a robust theoretical basis in selecting a distribution law, based on deduction rather than on trial-and-error procedures. Yet, the resulting maximum entropy distribution is not unique as it depends on the entropic form maximized and the constraints imposed. Here we use the Boltzmann-Gibbs-Shannon entropy and we propose a rationale for defining and selecting constraints. We suggest simple and general constrains that are suitable for positive, highly varying and asymmetric random variables, and lead to distributions consistent with geophysical processes. We define a generalization of the classical moments (the *p*-moments) which naturally leads to power-type distributions avoiding the use of generalized entropic measures.

2. Entropy measures

- Entropy as a concept dates back to the works of Rudolf Clausius in 1850 and of Ludwig Boltzmann around 1870 who gave entropy a statistical meaning and related it to statistical mechanics. Next, the concept of entropy was evolved by J. Willard Gibbs and Von Neumann in quantum mechanics, and was reintroduced in information theory by Claude Shannon in 1948.
- Information entropy is a purely probabilistic concept and is regarded as a measure of the uncertainty related to a random variable (RV).
- In literature there are more than twenty different entropy measures [1], proposed mainly as generalizations of Boltzmann-Gibbs-Shannon (BGS) entropy, which is the most famous and well justified entropy measure. The BGS entropy for a non-negative continuous RV X with density function $f_X(x)$ is defined as

$$S_{\mathrm{BGS}} = -\int_0^\infty f_X(x) \ln f_X(x) \mathrm{d}x \tag{1}$$

- A famous generalization, proposed by Rényi in 1961, is defined as

$$S_{\mathrm{R}} = \frac{1}{1-q} \ln \int_0^\infty f_X(x)^q \, \mathrm{d}x \tag{2}$$

- Another popular generalization, the Havrda-Charvat-Tsallis (HTC) entropy [2,3], is defined as

$$S_{\mathrm{HCT}} = \frac{1}{1-q} \left[\int_0^\infty f_X(x)^q \mathrm{d}x - 1\right] \tag{3}$$

- For $q \to 1$ both the Renyi and HCT entropies converge to the BGS entropy.

3. The principle of maximum entropy

- The principle of maximum entropy (PME), established by Edwin Jaynes [4,5], essentially relies in finding the most suitable probability distribution under the available information. According to Jaynes, the resulted maximum entropy distribution "is the least biased estimate possible on the given information…".
- Mathematically, the given information used in the principle of maximum entropy, is expressed as a set of constraints formed as expectations of functions $g_j(\ )$ of X, i.e.,

$$E\left[g_j(x)\right] = \int_0^\infty g_j(x)\, f_X(x)\mathrm{d}x = c_j, \quad j = 1, \dots, n \tag{4}$$

- The resulting maximum entropy distributions emerge by maximizing the selected form of entropy with constraints c_j, and with the additional constraint (to guarantee the legitimacy of the distribution)

$$\int_0^\infty f_X(x)\,\mathrm{d}x = 1 \tag{5}$$

- The general solution of the maximum entropy distributions resulting from the maximization of BGS entropy and the HCT entropy (accomplished by using the method of Lagrange multipliers) are, respectively,

$$f_X(x) = \exp[-\lambda_0 - \textstyle\sum_{j=1}^{n} \lambda_j\, g_j(x)] \tag{6}$$

$$f_X(x) = \left\{1 + (1-q)\left[\lambda_0 + \textstyle\sum_{j=1}^{n} \lambda_j\, g_j(x)\right]\right\}^{-1/(1-q)} \tag{7}$$

where λ_j, with j =0,…, n are the Lagrange multipliers linked to the constraints.

4. Selecting the constraints

- The choice of the imposed constraints is the most important and determinative part of the method as it defines uniquely the resulting maximum entropy distribution.
- Choosing constraints, however, is not trivial; theoretically, the expectation of any RV function can be used.
- Commonly, entropy maximization is done by assuming known mean and variance, which leads (a) to the Gaussian distribution in the BGS entropy case and (b) to a symmetric bell-shaped distribution with power-type tails in the HCT entropy case.
- So, how should we choose constraints?

i. Constraints should express our state of knowledge concerning a RV and should summarize all the available information from observations or from theoretical considerations.

ii. We can assume that some coarse features of the RV, e.g., the mean or the variance, are more likely to be preserved in the future than finer features, e.g., the kurtosis coefficient. Therefore, constraints should be simple and express features that are robust to estimate from the sample, and are likely to be preserved in the future.

iii. For some geophysical processes we may know important prior characteristics of the underlying distribution that should be preserved, e.g., a J-shaped or bell-shaped distribution or a heavy- or light-tailed distribution. So, constraints also should be chosen based on the suitability of the resulting distribution regarding the empirical evidence.

5. The generalized power function and p-moments

- Here, we aim to define some simple and general constraints, to use with the BGS entropy, that lead to suitable probability distributions for geophysical processes.
- Many entropy generalizations have emerged to explain empirically detected deviations from exponential type distributions that arise from the BGS entropy using moment constraints. Yet, generalized entropy measures have been criticized for lacking theoretical consistency and for being arbitrary.
- Here, we generalize the important notion of moments inspired by the limiting definition of the exponential function. We first define the generalized power function

$$x_p^q = \ln(1 + p\, x^q)/p \tag{8}$$

which for $p \to 0$ becomes the familiar power function x^q. Thus, we can define a generalization of the classical moments, which we name p-moments of order q as

$$m_q^p = E(X_p^q) = \frac{1}{p}\int_0^\infty \ln(1 + p\, x^q)\ f_X(x)\mathrm{d}x \tag{9}$$

Clearly, for $p \to 0$, p-moments are identical to classical moments, i.e., $m_q^0 = m_q \equiv E(X^q)$.

- We believe that the following rationale supports the use of p-moments:
   i. Generalized entropy measures have been successfully used; why not p-moments with the standard definition of entropy?
   ii. Maximization of the BGS entropy using p-moments leads to flexible power-type distributions (including the Pareto and Tsallis distributions for q = 1 and q = 2, respectively).
   iii. p-moments are simple and, for p = 0, become identical to the ordinary moments.
   iv. They exhibit similar properties with the $\ln x$ function, and thus are suitable for positively skewed RVs; additionally, compared to $E(\ln x)$ they are always positive.

6. The expectation of ln X as a constraint

- The generalized power function, as defined in (8) involves the logarithmic function, so, it would be useful to investigate also the logarithmic function as a constraint.
- The geometric mean μ_{G} given by

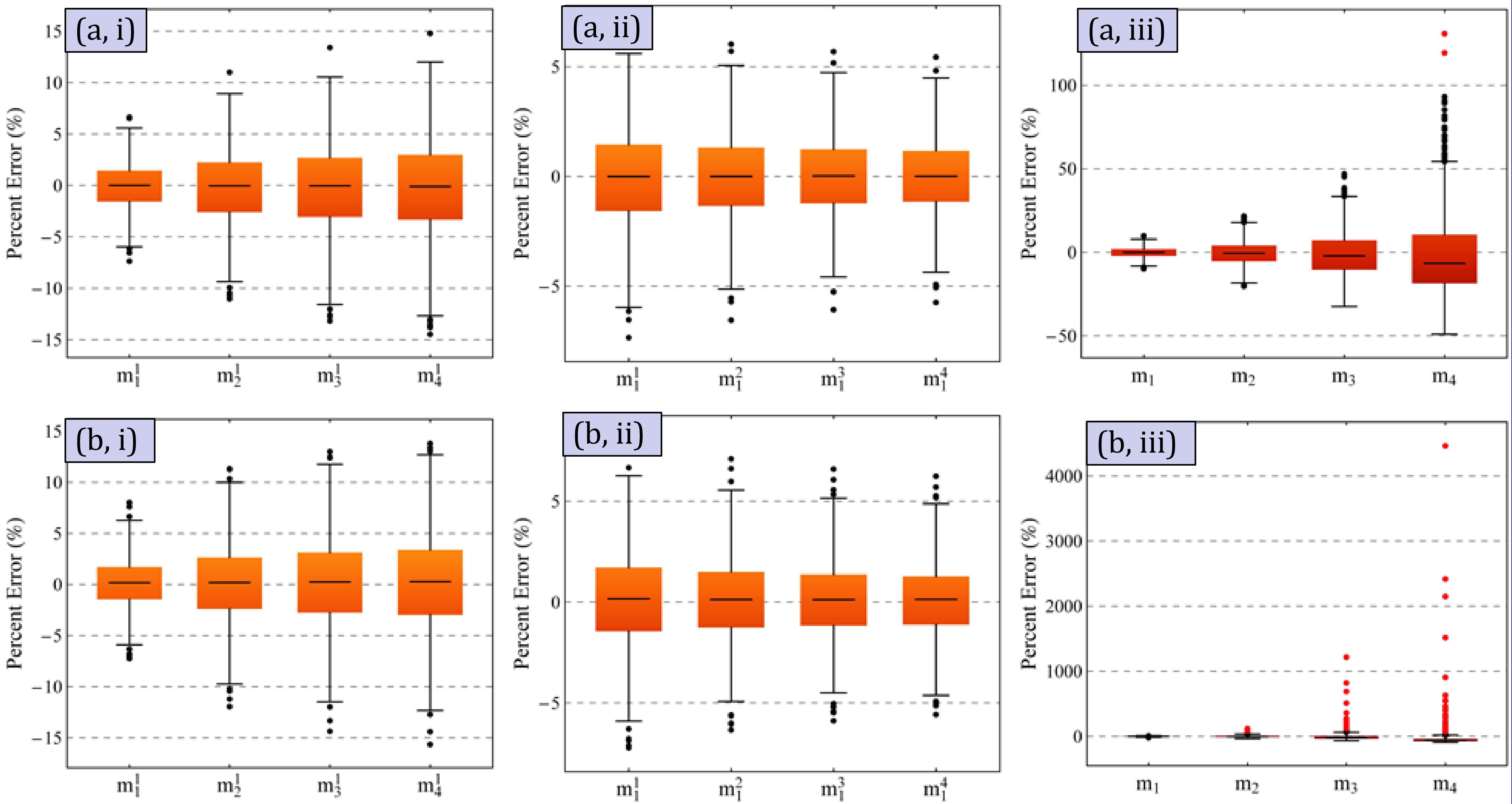
$$\mu_G = \left(\prod_{i=1}^{n} x_i\right)^{1/n} = \exp\left(\frac{1}{n}\sum_{i=1}^{n} \ln x_i\right) = \exp\left(\overline{\ln x}\right) \tag{10}$$

  is measure of central tendency, with the convenient property for geophysical processes to be defined only for positive values. This gives an intuitively meaning to formulate the expectation of ln X as a constraint

$$E(\ln X) = \ln \mu_{\mathrm{G}} \tag{11}$$

- Apart from its relationship to the geometric mean and its simplicity, the expectation of ln X, has some desired properties that make it an essential constraint for positively skewed RVs. Samples drawn from positively skewed, or even more, from heavy-tailed distributions, e.g., like those of daily rainfall, exhibit values that act like outliers and consequently strongly influence the sample moments, especially those of higher order. On the contrary, the function ln x applied to this kind of samples eliminates the influence of those “extreme” values and offers a very robust measure that is more likely to be preserved than the estimated sample moments. For this reason the logarithmic transformation is probably the most common transformation used in hydrology as it tends to normalize positively skewed data.

7. *p*-moments estimates *vs*. moments estimates



The figure depicts Monte Carlo results, i.e., we generated 1000 random samples (1000 values each) from each of the two following distribution (a) an Exponential with scale parameter $\beta = 1$ and (b) a Pareto type II with scale parameter $\beta = 1$ with asymptotic behavior $\sim x^{-5}$. We estimated the percent error defined as Percent Error % $= 100\ (sm_q^p - m_q^p)/m_q^p$, where sm_q^p is the sample estimate of m_q^p for the following three cases (i) for $p = 1$ and $q = 1 - 4$, (ii) for $q = 1$ and $p = 1 - 4$, and (iii) for the first four classical moment ($m_q^0 = m_q$). The percent error of the classical moments is orders of magnitude higher, if compared to that of *p*-moments.

8. Entropy maximization based on p-moments

The following table displays distributions (in terms of Lagrange multipliers λ_j) arising from the maximization of the BGS entropy and by imposing constraints, (a) classical moments m_q of various orders, (b) p-moments or various orders, and (c) combinations of moments or p-moments with the expectation of $\ln x$. In all cases classical moments produce exponential-type distributions while p-moments produce power-type distributions.

Constraints	Distribution Name	Density function
m_1	Exponential	$f_X(x) = \exp(-\lambda_0 - \lambda_1 x)$
m_2	Half-Normal	$f_X(x) = \exp(-\lambda_0 - \lambda_1 x^2)$
m_1 and m_2	Normal	$f_X(x) = \exp(-\lambda_0 - \lambda_1 x - \lambda_2 x^2)$
m_q	Generalized Exponential	$f_X(x) = \exp(-\lambda_0 - \lambda_1 x^q)$
$E(\ln x)$ and m_1	Gamma	$f_X(x) = x^{-\lambda_1} \exp(-\lambda_0 - \lambda_2 x)$
$E(\ln x)$ and m_q	Generalized Gamma	$f_X(x) = x^{-\lambda_1} \exp(-\lambda_0 - \lambda_2 x^q)$
m_1^p	Pareto type II	$f_X(x) = \exp(-\lambda_0)(1 + px)^{-\lambda_1/p}$
m_2^p	Tsallis	$f_X(x) = \exp(-\lambda_0)(1 + px^2)^{-\lambda_1/p}$
m_1^p and m_2^p	Not named	$f_X(x) = \exp(-\lambda_0)\left[(1 + px)^{\lambda_1}(1 + px^2)^{\lambda_2}\right]^{-1/p}$
$E(\ln x)$ and m_1^p	Beta of the second kind	$f_X(x) = \exp(-\lambda_0)\, x^{-\lambda_1}(1 + px)^{-\lambda_2/p}$
$E(\ln x)$ and m_q^p	Generalized Beta of the second kind	$f_X(x) = \exp(-\lambda_0)\, x^{-\lambda_1}(1 + px^q)^{-\lambda_2/p}$

9. The Generalized Beta of the second kind (GB2)

- The GB2 distribution, arises from the BGS entropy maximization by imposing as constraints the expectation of the ln x and the p-moment of arbitrary order q. The density function in terms of Lagrange multipliers is

$$f_X(x) = \exp[-\lambda_0 - \lambda_1 \ln x - \lambda_2 \ln(1 + p\,x^q)\,/p] \tag{12}$$

which after algebraic manipulations and parameter renaming can be written as

$$f_X(x) = \frac{\gamma_3}{\beta\,B(\gamma_1,\gamma_2)}\left(\frac{x}{\beta}\right)^{\gamma_1\ \gamma_3-1}\left[1+\left(\frac{x}{\beta}\right)^{\gamma_3}\right]^{-(\gamma_1+\gamma_2)}, x \geq 0 \tag{13}$$

corresponding to the distribution function

$$F_X(x) = B_z(\gamma_1,\gamma_2)/B(\gamma_1,\gamma_2), \text{where } z = [1 + (x/\beta)^{-\gamma_3}]^{-1} \tag{14}$$

where $B(\,,\,)$ and $B_z(\,,\,)$ are the Beta and the incomplete Beta functions, respectively.

- The GB2 distribution is probably one of the most versatile power-type distributions in literature and has been rediscovered many times under different names and parameterizations. It seems that Milke and Johnson [6] were the first who formed this distribution, and proposed it for describing hydrological and meteorological variables.
- The GB2 distribution is an extremely flexible four-parameter distribution comprising one scale parameter β, and three shape parameters $\gamma_1, \gamma_2, \gamma_3$, which allow the distribution to form innumerable different shapes. Many of the well-known distributions are special or limiting cases of the GB2 distribution (see e.g. [7,8]).

10. A simple special case of GB2 distribution

- While the GB2 distribution is extremely flexible, it is also very complicated and thus not easy to handle. Nevertheless, among the several three-parameter special cases that can be constructed, a simple case with analytical distribution function is obtained by setting $\gamma_1 = 1$ in the density function of GB2. After some trivial algebraic manipulations and parameter renaming we obtain a distribution known as the Burr type XII, introduced by Burr in 1942 in the framework of distribution system similar to Pearson's. Its density function is

$$f_X(x) = \frac{1}{\beta}\left(\frac{x}{\beta}\right)^{\gamma_1 - 1}\left[1 + \gamma_2\left(\frac{x}{\beta}\right)^{\gamma_1}\right]^{-\frac{1}{\gamma_1\gamma_2}-1}, x \geq 0 \qquad (15)$$

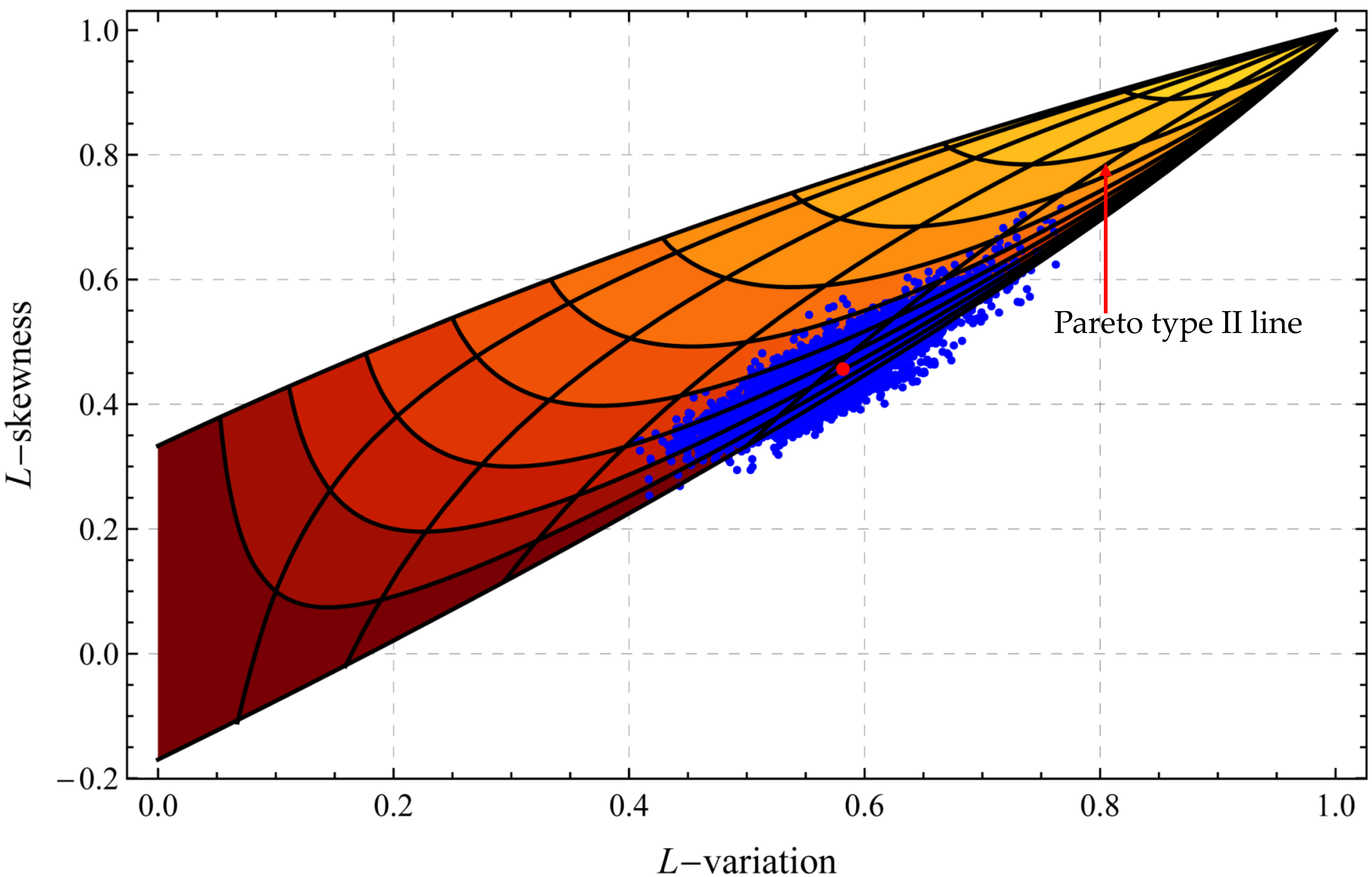
which corresponds to the distribution function

$$F_X(x) = 1 - \left[1 + \gamma_2\left(\frac{x}{\beta}\right)^{\gamma_1}\right]^{-\frac{1}{\gamma_1\gamma_2}} \qquad (16)$$

- The Burr type XII distribution is a flexible power-type distribution that comprises the scale parameter β and the shape parameters γ_1 and γ_2. The distribution is found in literature in a different form than the one given here (see e.g. [9]). However, we prefer this expression as it constitutes a kind of a generalization of the familiar Weibull distribution (for $\gamma_2 \to 0$). Additionally, the asymptotic behavior of the right tail is solely controlled by the parameter γ_2, while the parameter γ_1 controls the left tail. Particularly, for $0 < \gamma_1 < 1$ the distribution is J-shaped, for $\gamma_1 > 1$ is bell-shaped while for $\gamma_1 = 1$ it degenerates to the familiar Pareto type II distribution.

11. Application to daily rainfall

To test the applicability of the above theoretical framework, we used a large data set of daily rainfall records, i.e., the Global Historical Climatology Network-Daily database (http://www.ncdc.noaa.gov/oa/climate/ghcn-daily) which includes data recorded at over 40 000 stations worldwide. Many of those records, however, are too short in length, have missing data, or, contain data suspect in terms of quality. Thus, we selected for analysis only those records fulfilling the following criteria: (a) record length greater or equal than 50 years, (b) missing data less than 10% and, (c) data with quality flags less than 0.1% (see the above web site for details about flags). The selected subset includes 11 697 daily rainfall records. In the analysis performed [10], we tested the suitability of the Burr type XII distribution based on *L*-moments ratio plots.



The figure depicts the theoretical area covered by the Burr type XII distribution in an *L*-skewness *vs.* *L*-variation plot. The blue dots represent the corresponding sample *L*-statistics of the 11 697 daily rainfall time series, while the red dot depicts their average value.

Most of the sample points, i.e., 89.5%, lie within the distribution's area. Among those points, 63.4% lies below the Pareto line and 36.6% above, corresponding thus to J-shaped and bell-shaped distributions, respectively.

Finally, the right tail of the distribution corresponding to the average sample point behaves asymptotically as $x^{-4.8}$.

12. Summary and conclusions

- The principle of maximum entropy offers a theoretical basis to derive probability distributions based on the available information. Nevertheless, the available information is expressed as constraints imposed in the maximization of entropy without any limitation on their form.
- Here, we argue that the selected constraints should (a) summarize all the available information either from observations or from theoretical considerations, and (b) be simple and express features that are robust to estimate from the sample and are likely to be preserved in the future.
- We introduce a generalization of classical moments, the p-moments and justify their use as constraints, especially for highly skewed and varying random variables.
- We study the expectation of $\ln X$ as a constraint and we justify its use as suitable for geophysical processes.
- A Monte Carlo study showed that the sample estimates of p-moments are more robust, if compared to that of classical moments.
- Optimization of the BGS entropy using p-moments as constraints, naturally leads to power-type distributions thus avoiding generalized entropy measures.
- Optimization of the BGS entropy, with constraints the p-moment order q and the expectation of $\ln X$ leads to a very flexible, power-type distribution, known as the Generalized Beta of the second kind (GB2).
- An empirical analysis performed to 11 697 daily rainfall records worldwide, showed that the Burr type XII distribution (a special case of the GB2 distribution) is a very good probabilistic model for positive daily rainfall.

References

[1] Esteban MD, Morales D. A summary on entropy statistics. Kybernetika. 1995;31(4):337-346.
[2] Havrda J, Charvát F. Concept of structural a-entropy. Kybernetika. 1967;3:30–35.
[3] Tsallis C. Possible generalization of Boltzmann-Gibbs statistics. Journal of Statistical Physics. 1988;52(1):479-487.
[4] Jaynes ET. Information Theory and Statistical Mechanics. Phys. Rev. 1957;106(4):620.
[5] Jaynes ET. Information Theory and Statistical Mechanics. II. Phys. Rev. 1957;108(2):171.
[6] Mielke Jr PW, Johnson ES. Some generalized beta distributions of the second kind having desirable application features in hydrology and meteorology. Water Resources Research. 1974;10(2):223–226.
[7] Kleiber C, Kotz S. Statistical size distributions in economics and actuarial sciences. Wiley-Interscience; 2003.
[8] McDonald JB, Xu YJ. A generalization of the beta distribution with applications. Journal of Econometrics. 1995;66(1-2):133-152.
[9] Tadikamalla PR. A Look at the Burr and Related Distributions. International Statistical Review / Revue Internationale de Statistique. 1980;48(3):337-344.
[10] Papalexiou S.M., Koutsoyiannis D. An entropy based model of the probability distribution of daily rainfall [tentative title], Advances in Water Resources [in review] 2011.